ARI Research Note 90-31

Retrieval of Knowledge Through Algorithmic Decomposition

Sarah Lichtenstein

Perceptronics, Inc.

Basic Research
Michael Kaplan, Director

June 1990

United States Army
Research Institute for the Behavioral and Social Sciences

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency Under the Jurisdiction of the Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Technical Director

JON W. BLADES
COL, IN
Commanding

Research accomplished under contract for the Department of the Army

Perceptronics, Inc.

Technical review by

Nehama Babin

Accession For	
NTIS GRA&I	X
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

NOTICES

DISTRIBUTION: This report has been cleared for release to the Defense Technical Information Center (DTIC) to comply with regulatory requirements. It has been given no primary distribution other than to DTIC and will be available only through DTIC or the National Technical Information Service (NTIS).

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The views, opinions, and findings in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other authorized documents.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS --
2a. SECURITY CLASSIFICATION AUTHORITY -- 2b. DECLASSIFICATION / DOWNGRADING SCHEDULE --	3. DISTRIBUTION / AVAILABILITY OF REPORT Available for public release; distribution is unlimited.
4. PERFORMING ORGANIZATION REPORT NUMBER(S) Perceptronics PDRIR-1129-9-87	5. MONITORING ORGANIZATION REPORT NUMBER(S) ARI Research Note 90-31
6a. NAME OF PERFORMING ORGANIZATION Perceptronics, Inc. 6b. OFFICE SYMBOL (If applicable) --	7a. NAME OF MONITORING ORGANIZATION U.S. Army Research Institute
6c. ADDRESS (City, State, and ZIP Code) 6271 Variel Avenue Woodland Hills, CA 91367	7b. ADDRESS (City, State, and ZIP Code) 5001 Eisenhower Avenue Alexandria, VA 22333-5600
8a. NAME OF FUNDING / SPONSORING ORGANIZATION U.S. Army Research Institute for the Behavioral and Social Sciences 8b. OFFICE SYMBOL (If applicable) --	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER MDA903-83-C-0198
8c. ADDRESS (City, State, and ZIP Code) 5001 Eisenhower Avenue Alexandria, VA 22333-5600	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 61102B PROJECT NO. 74F TASK NO. WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)

Retrieval of Knowledge Through Algorithmic Decomposition

12. PERSONAL AUTHOR(S)

Lichtenstein, Sarah

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM 83/04 TO 87/09	1990, June	12

16. SUPPLEMENTARY NOTATION

Contracting Officer's Representative, Michael Drillings

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	~~Knowledge retrieval~~,
			Decomposed quantitative judgments (KR)

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This report summarizes the results of a series of studies to evaluate the effectiveness of algorithms for estimating unknown quantities. The results of the first experiment show that, as the structure of the aid increases, the subjects' performance improves in terms of both accuracy and consistency across subjects. The second experiment, however, suggests that algorithms used without understanding are of limited help. Further experiments suggest substantial problems in designing decision aids based on algorithmic decomposition, not because the principles of creating algorithms are hard to learn but because the users may be misled by their own misinformation and lack of arithmetic skills. These findings lead us to believe that such decision aids should be used in situations where algorithms can be carefully and deliberately designed and computational aids are also available. Possible approaches are suggested for the design of an expert system using algorithmic decomposition. Keywords:

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Michael Drillings	22b. TELEPHONE (Include Area Code) (202) 274-8722 22c. OFFICE SYMBOL OBR

DD Form 1473, JUN 86 Previous editions are obsolete. SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

RETRIEVAL OF KNOWLEDGE THROUGH ALGORITHMIC DECOMPOSITION

Decision makers often need to know the value of a particular quantity, such as "How much money did our family spend on food last year?" or "What is the Soviet troop strength in Cuba?" Sometimes the quantitative value is readily obtainable from a computer database, encyclopedia, reference source, or expert. In other cases, one may have read or been told the value so that it can be accurately and confidently retrieved from memory when it is needed.

There are some quantities, however, that may not exist in any reference sources or may be available only in sources that are difficult to locate, prohibitively costly to obtain, take too long to use, or contain only vague and partial knowledge. Faced with this situation, the best one can do is make an estimate of the needed quantity based on the information resources at one's disposal. In principle, those resources could extend to decision support systems, computer databases, information libraries, and the judgments of experts. In practice, however, estimates are often based on whatever relevant knowledge an estimator is able to obtain.

Even when a computer database is available, its contents may contain information related to the quantity called for, but not the quantity itself. Indeed, making full use of information systems entails understanding both the capacity of the system to provide direct answers to questions and its capacity to provide relevant information for questions it cannot answer directly. Effective querying of the system in the latter case requires a careful structuring of the user's information requirements, the absence of which can lead to inefficient use of computer resources, incomplete information retrieval, or erroneous results. How well one is able to exploit computerized information resources in such situations is in part dependent upon one's ability to structure whatever bits of information the system can provide in a form that is meaningful for the task of estimating the quantity in question. While the substantive contents of that structuring are potentially available from an information system, the form of the structuring is usually left to the idiosyncracies of the individual user. Studying the behavioral properties of various methods of structuring information helps shed light on the relative performance that might be expected from individuals when they are provided specific guidance on how to approach organizing quantitative estimation problems.

Approaches to structuring knowledge retrieval for the sake of estimating an uncertain quantity can vary in form and elaboration. The simplest approach is to consider what one knows about a quantity of interest and intuitively divine an estimate that seems reasonable in light of whatever knowledge comes to mind. This wholistic approach to estimation relies heavily on the power of unaided human cognition. To its advantage, it is inexpensive, portable, and represents the way in which people routinely deal with estimation problems, thus achieving a high degree of psychological compatibility. To its disadvantage, memory could prove to be too impoverished a resource on which to base an estimate and may provide no indication of which information might prove useful. Even when seemingly useful knowledge is retrieved, there may be no way of knowing how to combine disparate pieces of knowledge into a global estimate. Furthermore, research on the psychology of human judgment has repeatedly demonstrated that simplifying cognitive strategies can lead to systematic judgmental biases.

An alternative to the wholistic, intuitive approach is analysis or decomposition. This involves breaking up or decomposing a problem into a series of sub-problems or components, each of which can be understood more easily and operated on separately. The components are then assembled according to a prescribed set of combination rules to yield a solution, estimate, or prediction. Decomposition is a divide-and-conquer approach that assumes the components of a problem to be more understandable and tractable than the undecomposed problem.

A variant of the decompositional approach is the algorithm. An algorithm is a series of steps or operations that, when sequentially applied, produce a solution to a problem. Essentially, algorithms work by providing an unambiguous procedure for solving problems. They help structure what is known about a problem, point out what is not known, and specify the rules by which information should be combined.

The research summarized here was designed to explore the effectiveness of algorithms as an aid for estimating unknown quantities. In order to evaluate effectiveness, we have used, in our research, quantities for which we have accurate estimates, gained from almanacs and other such sources. However, we have used quantities for which people are unlikely to have ready estimates, such as the number of pounds of potato chips consumed annually in the United States.

In our first study (MacGregor, Lichtenstein & Slovic, in press), we selected sixteen questions about uncertain quantities like "How many people are employed by hospitals in the U.S.?" These sixteen questions were presented under five different aiding conditions. The most-aided group of subjects were given complete algorithms, created by the experimenters, for each question (for example, Average number of hospitals per state times number of states times average number of employees per hospital equals number of people employed by hospitals). These subjects were asked to make estimates for all the parts of the algorithm and to combine the parts as indicated to arrive at an estimate of the desired quantity. The second group of subjects were given the same algorithm without any indication of how to combine the parts. They were asked to make an estimate for each component and then estimate the desired quantity. The third group of subjects were asked to list components or factors they thought were relevant to the question, to make an estimate of each item on their list, and finally, to estimate the desired quantity. The fourth group of subjects were asked to make the same sort of list, but they were not asked to make estimates of each item before making an estimate of the desired quantity. The fifth group received no aid.

The results of this experiment showed that with increasing structure of the aid, the subjects showed improved performance in terms of both accuracy and consistency across subjects.

For factual questions like the ones used in this experiment, a decision maker would not normally have an algorithm readily available. In our second experiment (Lichtenstein & MacGregor, 1984), we explored the effectiveness of algorithmic aids on a class of numerical problems for which an algorithm is known. These problems are problems involving the combining of two pieces of probabilistic information. They are presented as story problems, like the following:

A light bulb factory uses a scanning device which is supposed to put a mark on each defective bulb it spots in the assembly line. Eighty-five percent (85%) of the light bulbs on the line are OK; the remaining 15% are defective.

The scanning device is known to be accurate in 80% of the decisions, regardless of whether the bulb is acually OK or actually defective. That is, when a bulb is good, the scanner correctly identifies it as good 80% of the time. When a bulb is defective, the scanner correctly marks it as defective 80% of the time.

Suppose someone selects one of the light bulbs from the line at random and gives it to the scanner. The scanner marks this bulb as defective.

What is the probability that this bulb is really defective?

We prepared two such problems, the light bulb problem shown above and a structurally identical problem involving a screening test for the reading disability, dyslexia, which has a false alarm rate and a false negative rate both equal to 5%. One of these problems was presented to three groups of subjects. One group received no aid. A second group were given an algorithm to solve the problem but no explanation as to why the algorithm represented the correct solution. The third group received a lengthy tutorial explaining how to solve such problems. After several intervening, unrelated tasks, all subjects were given the second problem, without any indication that it was related to the previous task and without any aid.

The results showed that the group given the unexplained algorithm showed better performance than did the control group on the first problem but that this improvement did not extend to the second problem. In contrast, the subjects given the lengthy tutorial showed improvement for both the first and second problems. These results suggest that algorithms used without understanding are of limited help.

Most of the uncertain quantities that decision makers face arise without the availability of an algorithm provided by someone else. Usually, decision makers must create their own algorithms. The effectiveness of algorithms in such situations was explored in two reports. For both studies, the subjects were given questions about uncertain quantities and answers to those questions. The subjects were assured that the answer provided by the experimenters was a wrong answer; the subjects' task was to decide whether the given answer was too high or too low.

In the first of these two studies (Lichtenstein, MacGregor & Slovic, 1987) we compared two types of instructions in how to create algorithms. In one case we told the subjects to momentarily disregard our answer and to build, from facts they knew or could estimate, an algorithm that would produce an estimate. Then they should compare their estimate with our answer to decide whether our answer was too high or too low. In the other case we told subjects to start with our answer and decompose it, using facts they knew or could estimate, to arrive at an implication of our answer that they could directly judge, to decide whether our answer was too high or two low.

The results of this study showed that, although subjects could create algorithms most of the time, there was only modest improvement in accuracy due to creating algorithms. There was no difference in accuracy as a function of the two types of instruction, although subjects more often used the former method than the latter. Further analyses showed that subjects were often misled by their own misinformation, such as believing that the population of the U.S. exceeds three billion people.

The second study about creating algorithms (Lichtenstein & Weathers, 1987) explored the effectiveness of asking people to create an algorithm and to use it twice, once to arrive at a reasonably low estimate of the desired quantity and again to arrive at a reasonably high estimate. The results of this study showed no significant improvement in accuracy due to this method. When subjects gave two estimates, both of which were larger than the given answer, they naturally concluded that our answer was too low. This led to excellent accuracy, of course, when our answer was indeed too low. But the subjects sometimes gave two estimates that were both larger than the given answer even when our answer was in fact too high. This situation happened often enough that the overall performance was no better than performance without algorithms.

Two further reports looked in depth at the difficulties subjects have in creating algorithms. The first (Lichtenstein & MacGregor, 1987) gave a detailed analysis of the types of arithmetic errors made by our subjects. Only 22% of our subjects were able to complete, without arithmetic or copying errors, a straightforward 13-step algorithm (one that required no estimation on the part of subjects). Although this result cannot be generalized to the population as a whole, it does suggest that builders of decision aids exercise great caution in assuming that their users have even the most basic, elementary arithmetic skills.

A similar cautionary tale can be drawn from the next report (Lichtenstein, 1987), which explored subjects' knowledge of ordinary facts. For example, only about 40% of the subjects, students at a state university, estimated the population of the U. S. as being between 200 million and 300 million. Only 33% could correctly state how many feet there are in a mile, although 78% knew how many cups are in a quart.

Taken together, these results suggest substantial problems in designing decision aids based on algorithmic decomposition, not because the principles of creating algorithms are hard to learn but because the users may be too often mislead by their own misinformation and their lack of arithmetic skills. This leads us to believe that the use of such decision aids might best be limited to situations in which experts can carefully and deliberately design algorithms for use in situations in which computational aids are also available. In our final paper for this project (Lichtenstein & Slovic, 1987), therefore, we discussed two possible approaches to the design of an expert system using algorithmic decomposition. To contrast the two approaches, one arising from the field of artificial intelligence, the other, from decision analysis, we took as our example the design of an expert system to predict dangerousness.

References

Lichtenstein, S. (1987). Factual knowledge of Oregon college students (Report 87-4). Eugene, OR: Decision Research.

Lichtenstein, S., & MacGregor, D. (1984). Structuring as an aid to performance in base-rate problems (Report No. 86-14). Eugene, OR: Decision Research.

Lichtenstein, S., & MacGregor, D. (1987). Arithmetic skills in using algorithms (Report No. 87-7). Eugene, OR: Decision Research.

Lichtenstein, S., MacGregor, D., & Slovic, P. (1987). Creating algorithms as an aid to judgment (Report No. 87-5). Eugene, OR: Decision Research.

Lichtenstein, S., & Slovic, P. (1987). Decomposition strategies for eliciting expert knowledge: Judgments of dangerousness (Report No. 87-8). Eugene, OR: Decision Research.

Lichtenstein, S., & Weathers, A. G. (1987). Creating algorithms as an aid to judgment: Part II (Report No. 87-6). Eugene, OR: Decision Research.

MacGregor, D., Lichtenstein, S., & Slovic, P. (In press). Structuring knowledge retrieval: An analysis of decomposed quantitative judgments. Organizational Behavior and Human Decision Processes.